二次元的散らばりを表わす統計量の分布

樋 口 伊 佐 夫

(1953 年 8 月 受付)

The Distribution of a Statistic Which Represents 2-Dimensional Dispersion

Isao HIGUCHI

Consider a bivariate distribution. Let σ_x and σ_y be the standard deviations of the variates, and ρ be their coefficient of correlation.

As Mr. C. HAYASHI says, the quantity $\sigma_x\sigma_y\sqrt{1-\rho^2}$ represents 2-dimensional dispersion of the distribution.

Concerning this quantity, which, we think, may also be available for some physiological study according to the theory of Dr. M. OSHIMA of Lab. of Labor Physiology, we get a mathematical result ——

Let (x_1, y_1), ····, (x_n, y_n) be a random sample of size n from a bivariate normal population, with σ_x, σ_y and ρ as its parameters. Let s_x, s_y and r be the maximum liklihood estimators of σ_x, σ_y and ρ respectively. Then $2n\, s_x s_y\sqrt{1-r^2}/\{\sigma_x\sigma_y\sqrt{1-\rho^2}\}$ is distributed according to the chi-square law with degree of freedom $2n-4$.

Institute of Statistical Mathematics

一次元分布の散らばりの度合を表わすのに分散又は標準偏差が適当であることと同様に，二次元分布の拡がり (extension) 或いは散らばり (dispersion) を表わすには

$$\sigma_x\sigma_y\sqrt{1-\rho^2}$$

なる量が適当であることは言うまでもない．(ここで σ_x, σ_y は夫々第一第二の変量の標準偏差，ρ は二つの変量の間の相関係数)

林知己夫氏は数量化の問題でこれをデータの classification などに用うべきものとして着目されたが，我々は Flicker 疲労測定のデータ解析の際に古典的な方法でつかつてみたいと考えた．そこで標本標準偏差 s_x, s_y, 標本相関係数 r を用いてつくつた統計量

$$s_xs_y\sqrt{1-r^2}$$

の統計的変動の様子を知ることは，いろんな意味で都合がよいと思う．該統計量の分布を母集団分布が二変数正規型の場合について計算してみた．正規母集団仮定をおくことは直接応用には制限を免れないが，いろいろな問題に対して一応の目安にはなると思う．

定理: (x_1, y_1), (x_2, y_2), ····, (x_n, y_n) を確率密度が

$$\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\exp\left[-\frac{1}{1-\rho^2}\left\{\frac{(x-\mu_x)^2}{2\sigma_x^2}-\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}+\frac{(x-\mu_y)^2}{2\sigma_y^2}\right\}\right]dxdy \tag{1}$$

なる分布をもつ母集団からとられた大きさ n の標本とし

$$\bar{x}=\frac{1}{n}\sum_1^n x_i \qquad s_x=\sqrt{\frac{1}{n}\sum_1^n(x_i-\bar{x})^2}$$

$$\bar{y}=\frac{1}{n}\sum_1^n y_i \qquad s_y=\sqrt{\frac{1}{n}\sum_1^n(y_i-\bar{y})^2}$$

$$r=\frac{\frac{1}{n}\sum_1^n(x_i-\bar{x})(y_i-\bar{y})}{s_x s_y}$$

とすると，　$s^2=s_x s_y\sqrt{1-r^2}$

の分布の密度は

$$\frac{1}{\Gamma(n-2)}\left(\frac{n}{\sigma^2}\right)^{n-2}(s^2)^{n-3}\exp\left\{-\frac{ns^2}{\sigma^2}\right\}d(s^2)\quad(s^2\geqq 0) \tag{2}$$

で与えられる．此処で σ^2 は $\sigma_x\sigma_y\sqrt{1-\rho^2}$ を表わす．従つて

系：$2ns^2/\sigma^2$ すなわち $2ns_xs_y\sqrt{1-r^2}/\{\sigma_x\sigma_y\sqrt{1-\rho^2}\}$は自由度 $2n-4$ の χ^2 分布に従う．

このことから二次元的ひろがり σ^2 に関する信頼区間や検定をつくることが出来る．

定理の証明

まづ

$$\Phi\equiv\begin{pmatrix}a & c\\ c & b\end{pmatrix},\qquad X\equiv\begin{pmatrix}x & z\\ z & y\end{pmatrix}$$

を夫々常数及び変量の二次の positive matrix とし，$\det\Phi=ab-c^2$，$\det X=xy-z^2$ を夫々 A 及び λ であらわすことにする．x, y, z は同時密度が $x\geqq 0$ $y\geqq 0$ 且 $\lambda\geqq 0$ なる範囲で

$$\frac{A^{\frac{m}{2}}\lambda^{\frac{m-3}{2}}}{2^m\sqrt{\pi}\,\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{m-1}{2}\right)}\exp\left\{-\frac{1}{2}(ax+by+2cz)\right\}dx\,dy\,dz$$

他では 0 なる Wishart 分布 $W_{m,2}(\Phi)$ に従うものとする．まづこのときの λ の分布を計算する．変数 (x,y,z) を (x,z,λ) に変換する．この変換は問題になる領域では明かに一対一であり，変換の Jacobian の絶対値は $1/x$ であるから x,z,λ の同時分布の密度は $x\geqq 0$ $\infty>z>-\infty$ $\lambda\geqq 0$ で

$$\frac{A^{\frac{m}{2}}\lambda^{\frac{m-3}{2}}}{2^m\sqrt{\pi}\,\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{m-1}{2}\right)x}\exp\left[-\frac{1}{2}\left\{\left(a-\frac{c^2}{b}\right)x+\frac{b\lambda}{x}\right\}\right]$$

$$\times\exp\left\{-\frac{1}{2}\cdot\frac{(z+cx/b)^2}{x/b}\right\}d\lambda\,dx\,dz$$

z を integral out して，x と λ の同時密度

$$\frac{A^{\frac{m}{2}}\lambda^{\frac{m-3}{2}}}{2^{\frac{2m-1}{2}}\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{m-1}{2}\right)}\frac{1}{\sqrt{bx}}\exp\left[-\frac{1}{2}\left\{\frac{A}{b}x+\frac{b\lambda}{x}\right\}\right]d\lambda\,dx$$

を得る．$x\geqq 0$, $b>0$, $A>0$ だから

$\sqrt{\dfrac{xA}{2b}}\equiv p$ とおき p と λ との同時密度を求めると

$$\frac{A^{\frac{m-1}{2}}\lambda^{\frac{m-3}{2}}}{2^{m-2}\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{m-1}{2}\right)}\exp\left\{-\left(p-\frac{\sqrt{A\lambda}}{2p}\right)^2\right\}\exp\{-\sqrt{A\lambda}\}\,d\lambda\,dp$$

ところで $\alpha>0$ に対しては一般に

$$\int_0^\infty \exp\left\{-\left(\frac{\alpha}{u}-u\right)^2\right\} du = \frac{\sqrt{\pi}}{2}$$

だから*

$$2^{m-2}\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{m-1}{2}\right)=\sqrt{\pi}\,\Gamma(m-1)$$

に注意して p を integral out すると λ の密度

$$\frac{1}{2\Gamma(m-1)}A^{\frac{m-1}{2}}\lambda^{\frac{m-3}{2}}\exp\{-\sqrt{A\lambda}\}\,d\lambda \tag{3}$$

を得る．$\sqrt{\lambda}=ns^2$ とすると s^2 の分布の密度は

$$\frac{n^{m-1}}{\Gamma(m-1)}A^{\frac{m-1}{2}}(s^2)^{m-1}\exp\{-\sqrt{A}\,ns^2\}\,d(s^2) \tag{4}$$

となる．ところで正規母集団（1）からの大きさ n の標本を用いてつくつた ns_x^2, ns_y^2 及び ns_xs_yr の同時分布は

$$\Phi_0=\frac{1}{1-\rho}\begin{pmatrix}\dfrac{1}{\sigma_x^2} & \dfrac{-\rho}{\sigma_x\sigma_y}\\ \dfrac{-\rho}{\sigma_x\sigma_y} & \dfrac{1}{\sigma_y^2}\end{pmatrix}$$

としたとき Wishart 分布 $W_{n-1,2}(\Phi_0)$ となるから，$n^2s_x^2s_y^2(1-r^2)$ は上記の分布（3）に於て $m=n-1$, $A=\det\Phi_0=\dfrac{1}{\sigma_x^2\sigma_y^2(1-\rho^2)}=\dfrac{1}{\sigma^4}$ とおいたものに従う．これから $s_xs_y\sqrt{1-r^2}$ の分布密度は（4）に於て m, A を同じくおきかえたもの即ち（2）になることがわかる．以上

つけたし： 散らばりを表わす量 s^2 の分布は母集団の散らばりを表わす σ^2 だけに依存するから二つの母集団からとられた二組の標本について，母集団 parameter $\sigma_x^{(i)}$, $\sigma_y^{(i)}$, $\rho^{(i)}(i=1,2)$ は異つていても $\sigma^{2(1)}=\sigma^{2(2)}$ かどうかの test が出来る．すなわち

$$F=\frac{n_1s_x^{(1)}s_y^{(1)}\sqrt{1-r^{(1)2}}/(m_2-2)}{n_2s_x^{(2)}s_y^{(2)}\sqrt{1-r^{(2)2}}/(n_2-2)}$$

が自由度 $2n_1-4$, $2n_2-4$ の F 分布に従うことを利用すればよい．

労動衛生学に於ける Flicker-value による疲労測定に際しては，作業前作業後の Flicker-value を夫々 x 及び y とすると，人間の集団に対して $(x, y-x)$ はほゞ二次元正規分布に従うと見做し得るようである．労働科学研究所大島正光博士**の理論によると，疲労の大きい集団に対しては $y-x$ の x に対する回帰係数が負の値で絶対値が大きくなるとともに回帰直線のまわりの $y-x$ の分散が小さくなる．通常回帰係数の有意差検定では回帰直線のまわりの分散は等しいと仮定している．ところで我々の取扱つたデータではその仮定は首肯し難い．分散の異りは荷重をつけることにより一応さけることが出来るが，σ_x の方は一定しているように思えるので $x, y-x$ の二次元正規分布に於ける σ^2 も疲労の度合をあらわす一つの measure として採用出来るものと思う．そこで体質の異る二つのグループに対して σ^2 が等しいかどうかの検定をつくつてみることを考えたのである．

或る実験データについては藤原長司君が計算を行つたが，実験企画そのものに疑問があり，また従来の古典的方法による我々の統計的推論の妥当性にも疑問があるので，ここでは具体的問題にはこれ以上立ち入らない．このことに関しては又別の機会に論じたいと思う．

（統計数理研究所）

---

* たとえば Gustav Doetsch: Theorie und Anwendung der Laplace-Transformation, Seite 25.

** 大島正光 他: 生体の機能の変動の見方について；（其一．作業前値の差の捨象について）労働科学 Vol, 28, No. 10（昭和 27 年）